Gaussian Processes and CNNs: The utility of Prediction Variance

Joshua Send [1]

Abstract

4-6 sentences TODO

1. Introduction

* Introductory paragraph * Problem with CNNs predictions * How GPs are bayesian predictors that give variance * This work combines ... by ... showing ... (or something) * ”Interpretable ML”

CNNs * Quick introduction to CNNs * How probabilities are calculated (Classification, softmax, alternatives discuss briefly) * Use as a feature extractor *

GPs * Quick introduction GPs * Bayesian, provide variance * Hopefully have higher variance when away from the space they are trained in * show 2D image of a fitted GP with variances for intuition * Covariance function/kernel steers behavior, different options are evaluated, can combine kernels but not some others * Classification versus regression

In general, how we can combine these * Using CNN as feature extractor, replacing Softmax with GP * What this paper will explore * MNIST, N-MNIST ((Basu et al., 2017)), Adverserial examples * Comparing CNN, GP with 2 different Kernels, two Hybrid models (ref to section where this decision is justified)

2. Related Work

* Variance from CNNs * GPDNN * Adverserial attacks, NN robustness to new examples (! TODO reading)

3. Implementation

* GPFlow, alternatives explored * Batching, different GP mechanisms to deal with O(N$\hat{3}$) scalability * Inducing points * Approximate trainining time for GP (=¿ discussion?) * Predict time for GP (=¿ discussion?)

* Keras MNIST *

* Train, test sizes across MNIST, NMNIST, Adverserial, and image sizes (28x28, grayscale) * Balanced datasets? (! TODO) * OPTIONAL: check performance across specific numbers?

* Hybrid model (=¿ own Section?)

4. MNIST

4.1. Accuracy of Various Models

* CNN performance * Show GP performance across various kernels * Many configurations were explored, results * White noise variance is not a big factor in performance * Performance is most correlated with * Explain decision to use Matern12 and Linear*Matern32 * inherits linear robustness to Blur/Low contrast * Inherits nice Matern32 properties for Hybridization * not that robust to adverserial... while Matern12 is

* footnote SVM as a example of another kernel method

4.2. Distribution of Variance

* Show CDF of incorrect, correct predictions * Discuss how this is useful

4.3. GP Variance

* Plot GP prediction probability versus confidence * Show it’s fully deterministic (TODO double check code to make sure it is) * Conclude variance is not actually extra information but useful for interpretation * Discuss interpretability, show some plots

4.4. Examining Misclassifications

* Plot CNN and GP mis-prediction probabilities * Discuss... conclusion is that CNN more confidently mis-predicts results? * Show examples that both fail, one fails * Discuss that nothing can be done when both fail - no extra information * Show overlap, non-overlap in misclassifications * Inspires Hybridization!

4.5. Hybridization

* Might be able to rescue these individual misclassifications * steered by variance! * Note that we give up interpretability here for higher accuracy in some cases, but we can notify the

[1] University of Cambridge. Correspondence to: Joshua Send <js2173@cam.ac.uk>.

users the uncertainty is higher because the models disagree! * some kernels better than other

* Describe both criteria, pros and cons of each * show some results with different criteria for some example where there's a significant difference between CNN and GP (Low contrast Linear*Matern32?) * Will show results from 'stronger@0.5' with Linear*Matern32 which performs best of all tested kernels at hybridization due to having the least overlapping misclassifications

Gains from Hybridization are usually not present

5. N-MNIST

5.1. White Noise + MNIST

(Sample image)

* Accuracy across, CNN, Matern12, Poly, Linear*Matern32, Hybridized * Distribution of correct, incorrect classification Variances for Linear*Matern32

5.2. Blurred MNIST

(Sample image) * Accuracy across, CNN, Matern12, Poly, Linear*Matern32, Hybridized * Distribution of correct, incorrect classification Variances for Linear*Matern32

5.3. White Noise and Low Contrast MNIST

(Sample image) * Accuracy across, CNN, Matern12, Poly, Linear*Matern32, Hybridized * Distribution of correct, incorrect classification Variances for Linear*Matern32

6. Adverserial Attacks

* Brief description of FSGM, that it uses the trained CNN to generate adverserial examples with some epsilon

* Accuracy across models as epsilon varies

* Distribution of correct, incorrect classification Variances for Linear*Matern32 for eps=0.2

Citations within the text should include the authors' last names and year. If the authors' names are included in the sentence, place only the year in parentheses, for example when referencing Arthur Samuel's pioneering work (**?**). Otherwise place the entire reference in parentheses with the authors and year separated by a comma (**?**). List multiple references separated by semicolons (**???**). Use the 'et al.' construct only for citations with three or more authors or after listing all authors to a publication in an earlier reference (**?**).

Authors should cite their own work in the third person in the initial version of their paper submitted for blind review. Please refer to Section **??** for detailed instructions on how to cite your own papers.

Use an unnumbered first-level section heading for the references, and use a hanging indent style, with the first line of the reference flush against the left margin and subsequent lines indented by 10 points. The references at the end of this document give examples for journal articles (**?**), conference publications (**?**), book chapters (**?**), books (**?**), edited volumes (**?**), technical reports (**?**), and dissertations (**?**).

Alphabetize references by the surnames of the first authors, with single author entries preceding multiple author entries. Order references for the same authors by year of publication, with the earliest first. Make sure that each reference includes all relevant information (e.g., page numbers).

6.1. Software and Data

Code, results, and this report can be found at:

https://github.com/flyingsilverfin/CNN_GP_MNIST

Please note that intermediate results are not saved but can be recomputed, and that some of the configurations are specific to the development machine.

Acknowledgements

Do not include acknowledgements in the initial version of the paper submitted for blind review.

If a paper is accepted, the final camera-ready version can (and probably should) include acknowledgements. In this case, please place such acknowledgements in an unnumbered section at the end of the paper. Typically, this will include thanks to reviewers who gave useful comments, to colleagues who contributed to the ideas, and to funding agencies and corporate sponsors that provided financial support.

References

Basu, Saikat, Karki, Manohar, Ganguly, Sangram, DiBiano, Robert, Mukhopadhyay, Supratik, Gayaka, Shreekant, Kannan, Rajgopal, and Nemani, Ramakrishna. Learning sparse feature representations using probabilistic quadtrees and deep belief nets. *Neural Processing Letters*, 45(3):855–867, 2017.

A. Do *not* have an appendix here

Do not put content after the references. Put anything that you might normally include after the references in a separate supplementary file.

We recommend that you build supplementary material in a separate document. If you must create one PDF and cut it up, please be careful to use a tool that doesn't alter the margins, and that doesn't aggressively rewrite the PDF file. pdftk usually works fine.

Please do not use Apple's preview to cut off supplementary material. In previous years it has altered margins, and created headaches at the camera-ready stage.